

Gender Difference in the Credibility Perception of Mobile Websites: A Mixed Method Approach

Kiemute Oyibo
University of Saskatchewan
Saskatoon, Canada
kiemute.oyibo@usask.ca

Yusuf Sahabi Ali
Ahmadu Bello University
Zaria, Nigeria
sahabiali@yahoo.com

Julita Vassileva
University of Saskatchewan
Saskatoon, Canada
jiv@cs.usask.ca

ABSTRACT

To persuade people to buy a product or service online they must be visually convinced and attracted to use the sales website. Thus, there is need to understand how different user groups perceive sites for better adaptation. A lot of research has shown that users' judgment of the credibility of a website is critical to its success. However, in the mobile domain, little has been done empirically to 1) investigate users' credibility perception of a website and 2) how it changes as the user interface (UI) design is systematically altered. This paper bridges this gap by carrying out sentiment and statistical analyses of users' perception of four systematically modified mobile websites among 285 subjects from North America, Africa and Asia. The results show that mobile website design affects the perception of its credibility, with 1) females being more critical and sensitive to UI changes than males; and 2) the grid-layout design preferred to list-layout design by both genders. The study contributes to the existing literature in two ways. First, it provides a concise model for understanding users' UI perceptions, expectations and gender differences. Second, it presents important findings that will enable a gender-based mobile web site adaptation.

CCS Concepts

• Human-centered computing → Human computer interaction (HCI) → HCI design and evaluation methods → User studies and user models.

Keywords

Mobile website; user model; user interface design; visual design; navigation; layout; credibility; gender difference; adaptation.

1. INTRODUCTION

The rapid growth in information technologies, especially the Internet and smartphones, has led to unprecedented opportunities for people to connect with one another, interact and trade. So far, many people have embraced the benefits of e-commerce. In the comfort of their home, or while on the move, they carry out online transactions through their laptops, tablets or smartphones. For example, they can order a product, book a plane ticket or hotel, etc., without having to go to the physical store, thereby saving time and money on transportation. Even for those who make their purchases in physical stores, research has shown that 70% of them use the retailer's websites and apps on their smartphones to seek online information on the products and services of interest prior to purchasing them [37]. This makes it more important than ever

before for mobile website owners, designers and advertisers to understand what exactly users expect from e-commerce websites in order to attract new customers and keep existing ones through data-driven design and adaptation [10]. Many e-commerce websites have been originally designed for use in a web browser and involve fairly complex workflows and when accessed on small-screen smart-phones, can create a confusing and untrustworthy experience for the user. Therefore, mobile e-commerce vendors need to provide a user-friendly interface for their customers to improve the user experience [23]. However, while there has been a number of empirical research on the influence of visual and navigation designs on web credibility, very few have been focused on the mobile domain and on the role gender plays [23]. Moreover, very few have been conducted among a mixed population, which cuts across diverse cultures, in order to arrive at more generalizable findings. For the most part, previous studies have focused on mainly Western and Asian demographics, often leaving out a continent like Africa, which happens to be one of the fastest growing mobile markets in the world today and a key player in the global mobile web [35]. To bridge this gap, we carried a mixed-method study of four systematically modified mobile websites among a mixed sample of 285 subjects from three continents (North America, Africa and Asia). In order to foster better mobile website design and adaptation, we investigate 1) how users perceive the various mobile websites in terms of aesthetics (e.g., color and images), usability (e.g., layout and spacing) and credibility; 2) how these perceptions vary as the visual and navigation characteristics are modified; and 3) the role gender plays in the various perceptions.

2. RELATED WORK

Researchers [4], [25], have found that inherent gender differences exist between males and females in the processing of information and across a wide range of cognitive tasks [32]. For example, males are known to perform better in spatial reasoning while females in verbal and linguistic activities [4], [14], [16], [24]. Females have also been found to be more visually discerning [15] and more accurate in decoding nonverbal cues [29] than males. In particular, in marketing and advertising, gender differences have been found to exist in the processing of advertising information [25], [9]; as a result, "*gender has been historically used as basis of market segmentation*" (p. 20) [32]. According to the selectivity model [25], females are *comprehensive* information processors who respond to subtle cues by considering a product's attributes both subjectively and objectively, while males are *selective* information processors who usually miss subtle cues because they process information heuristically. This was proven to be true by Arcand and Nantel [1], who carried out a study to investigate gender differences in search patterns and online task performance among 125 actual consumers. They found out that women spent significantly more time per page than men. Similarly, in the web domain, research [8], [11], [18], [30], [32] has shown that males and females perceive websites differently, with the former being more critical. In a study among 76 participants, Cyr and Bonanni [8], found that gender played a major role in the assessment of information and navigation design,

SAMPLE: Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

Conference '10, Month 1–2, 2010, City, State, Country.

Copyright 2010 ACM 1-58113-000-0/00/0010 ...\$15.00.

DOI: <http://dx.doi.org/10.1145/12345.67890>

with males being more satisfied than females. Ferebee [10] also found that males rated websites higher on credibility than females.

In the mobile domain, very few studies, regarding the perception of websites and gender difference, have been carried out [23], using a mixed method approach and a mixed sample. Li and Ye [19] carried out a study among 200 subjects and came up with structural equation model showing that design aesthetics indirectly impacted customer's trust. However, this study was based on a homogeneous population and did not look at the role gender plays in the perception of trust. Cyr et al. [5], in a similar study with 60 participants comprising 30 Canadians and 30 Chinese, found as well that design aesthetics indirectly influences loyalty to mobile websites, but could not find any significant influence of gender, culture or age. However, unlike our study, their sample size was small and did not include participants from Africa, which is currently one of the fastest growing markets in the world using the mobile web [35]. More recently, Lu and Rastrick [23] carried out a survey to investigate the influence of website design on the intention to adopt mobile commerce. They found that navigation design most significantly influenced users' perceived ease of use of mobile websites, and this was a more important factor for females than for males when deciding to use mobile commerce. However, their findings were based on quantitative analysis only, whereas ours adopted a mixed-method approach, focusing on the qualitative analysis, complimented and confirmed by quantitative results.

3. METHODOLOGY

In this section, we present our research design, instruments used in measuring constructs and the demographics of participants.

3.1 Research Design

The aim of our study is to investigate users' credibility perception, how it changes as the aesthetic and usability elements change and the role gender plays for better adaptation. So, we came up with what we called a "Mobile Web UI Transformation Framework" or "Action-Artifact (A2) Framework" to systematically modify the UI design of four hypothetical mobile webpages [26], adapted in 2014 from m.wakanow.com, mobile.united.com, mobile.utah.com and tourismwinnipeg.com. Fig. 1 shows the framework in a Cartesian coordinate system. The axes represent the actions (UI treatments) carried out in a clockwise direction to realize a new artifact (UI) in the next quadrant. We regard the UI pairs above and below the x-axis as low-level and high-level web designs respectively. Starting from the low-level group, we carry out a compound UI treatment (*make gray and add icon*) on A to produce B. Next, we carry out a simple UI treatment (*make unicolor*) on B to produce C. This UI transformation continues till we return to A from where we started. Finally, based on the four web UIs, we hypothesized as follows:

- H1:** Users' perception of credibility of mobile websites changes as the UI designs are modified.
- H2:** Users will be more concerned about visual than navigational design elements in judging the mobile websites.
- H3:** D will be judged as the best by both genders.
- H4:** A will be judged as the worst by both genders.
- H5:** Females will be more critical in their judgment of the mobile websites than males.

Our hypotheses, for the most part, were based on previous findings in the literature in the web domain, where most of the existing research has been focused. The first hypothesis (H1) was informed by the work of Robins and Holmes [28]. They found out that when the same web content was presented to users at different levels of

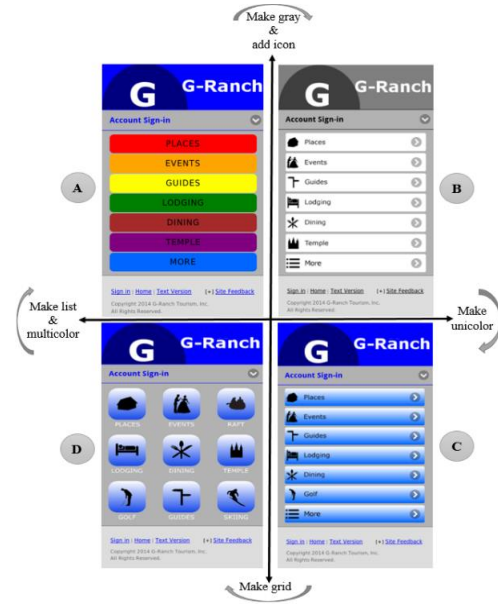


Figure 1. UI Transformation framework

aesthetic treatment, the one(s) with better aesthetic treatment performed better with respect to credibility assessment. The second hypothesis (H2) was informed by the work of Fogg [13], [12] and others [7], [19], [21], [20] on web credibility. They showed that, at the visceral level and for the most part, it is the perception of aesthetics, i.e., visual design, which influences users' judgment of website credibility. Thus, we believe that this would be true in the mobile domain as well despite the importance of usability, which, given the small-screen size of mobile devices, may make usability even more important. The third and fourth hypotheses (H3 and H4) stemmed from our judgement, as we view D and C, which belong to the high-level group, as more appealing UIs, while A and B, which belong to the low-level group, as less appealing UIs. So, between D and C, we speculated that the former (grid-based UI) will perform better than the latter (list-based UI), as the former appears to be more usable or convenient to use than the latter. We hypothesized that this better *usability* perception of D will impact the perception of the *visual design* and the *credibility* of the entire website as well due to the halo effect [33], thereby making it the best preferred. On the other hand, we surmised A will be judged as the worst because of the less professional choice and multiplicity of colors. Finally, the fifth hypothesis (H5) was based on gender-related findings in previous research. Using a homogeneous sample of 76 participants, Cyr and Bonanni [8] found out that significant gender differences exist in the way participants evaluated websites on the basis of design and satisfaction. Furthermore, based on a heterogeneous sample of 1156 subjects from 8 countries, Cyr et al. [6] found that men and women perceived the same websites differently based on a number of design characteristics, which included information design, navigation design, visual design, trust and satisfaction. Similarly, Flanagan and Metzger [11] found that there is a moderating effect in the way males and females evaluate website design and credibility. For example, Cyr and Bonanni [8], Flanagan and Metzger [11], and Ferebee [10] found that females rated websites less favorably than males. This was attributed to: 1) women are usually more critical in the judgement of things and information technology in particular [30], [18]; and 2) Most websites are designed to meet male rather than female preferences. As cited in [3], a study of UK websites found that 94% of the sites had a masculine orientation and 74% were designed by males.

The rest of this paper is organized as follows. Section 2 focuses on related work; Section 3 on the methodology of the study. Sections 3 and 4 present the results and discussion respectively. Finally Section 6 focuses on the conclusion.

3.2 Measures

Credibility perception was measured by using a combination of quantitative method (rating and ranking) and qualitative method (comments). First, participants were asked to rate each of the webpages on a Likert scale ranging from 1 to 7 and comment on what interested or annoyed them. Second, they were requested to rank them from 1 to 4. The four webpages were not presented to participants in any special order, e.g., from best to worst, or vice versa, as perceived by us the designers. Rather, they were presented out of order: C, A, B and D. We chose a single-item credibility rating scale because: 1) Bergkvist and Rossiter [2] have shown that “there is no difference in the predictive validity of the multiple-item and single-item measures” (p. 174); 2) the single-item has been used in a prior study [31]; and 3) to prevent participant fatigue.

3.3 Participants

The survey was approved by the University of Saskatchewan Research Ethics Board. Thereafter, it was posted on the university’s website and social network (Facebook) for anonymous participation. Also, invitation emails were sent to volunteer participants for a chance to participate. In order to appreciate participants for their time, they were given a chance to optionally enter for a draw to win one of our four gift cards worth \$50 each. The data gathering lasted for a period of six months. A total number of 300 subjects took part in the study. However, after cleaning, we were left with 285 valid participants, which include 149 (52.3%) males and 136 (47.7%) females. Table 1 shows the participants’ demographics. About 65.6% of the participants were between 18 and 24 years old, while the rest were older. Only about 66% of the participants provided comments, at least on one of the UIs (90 males and 87 females). About 45% of the participants had over ten years of internet experience, while 54% and 25.3% of them had high school and bachelor educational qualification. Further, the African, North American and Asian participants formed 54.7%, 33.3%, and 12.0% of the sample respectively.

4. RESULTS

In this section, we present the results of our analysis, which include comments word count, clustering of comment documents (files), word clouding of comment files, sentiment and statistical analyses.

4.1 Comments Word Count

We performed a word count on participants’ comments on all four UIs to find out which resonated the most with them and the gender differences. In total, we have eight comment documents (files), four for each gender. Table 2 shows the summary of the word count with stopwords removed. Overall, C and B elicited the highest and lowest number of words (709 and 451) respectively. A possible explanation for the former is that C was presented to participants first in the survey. Further, the female group (FG, 1,265) scored a higher word count than the male group (MG, 932) across all UIs, suggesting that the females were more stimulated to respond than the males given that the ratio of male to female commenters is roughly 1 to 1 (90 to 87), as shown in Table 1. It also confirms the theory [22] that females generally tend to be more verbal or talkative.

4.2 Comment files Clustering

We carried out K-Means clustering on the comments of participants in order to understand how their perceptions changed the UI is

transformed from one design to another. Figure 3 shows a principal component plot of three clusters. The first principal component represents the maximum possible variability (18%) in the eight comment files (for each artifact and gender - one document with all the comments), followed by the second principal component which accounts for 16% of the total variation. Cluster 1 indicates that males and females responded in a similar way to A with respect to the choice of words used. Similarly, Cluster 2 indicates males responded in a similar way to B, C and D, with more similarity existing between M_B and M_D due to their overlapping. Lastly, based on their proximity on the plot, Cluster 3 indicates that females responded in a similar way to B, C and D, with more similarity existing between F_B and F_D. We discuss the significance of the principal components in subsection 4.3.

Table 1. Sample Demographics (n=285)

Criterion	Group	Number	Percent
Gender	Male	149	52.3%
	Female	136	47.7%
Age	18-24	187	65.6%
	25-34	79	27.7%
	>44	19	6.7%
Continent	Africa	156	54.7%
	North America	95	33.3%
	Asia	34	12.0%
Country	Nigeria	147	51.6%
	Canada	92	32.3%
	China	9	3.2%
	Others	37	12.9%
Years on Internet	<10	127	44.6%
	>=10	158	55.4%
Educational Qualification	High School	154	54.0%
	Bachelor	72	25.3%
	Postgraduate	35	12.3%
	Others	24	8.4%
Commenter	Male	90	31.6%
	Female	87	30.5%

Table 2. Comment word count with stopwords removed

Webpage	Male	Female	Global
A	211	344	555
B	201	250	451
C	324	385	709
D	196	286	482
Total	932	1,265	2,197

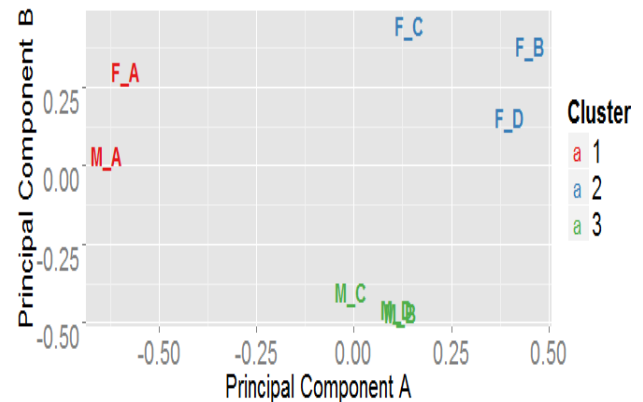


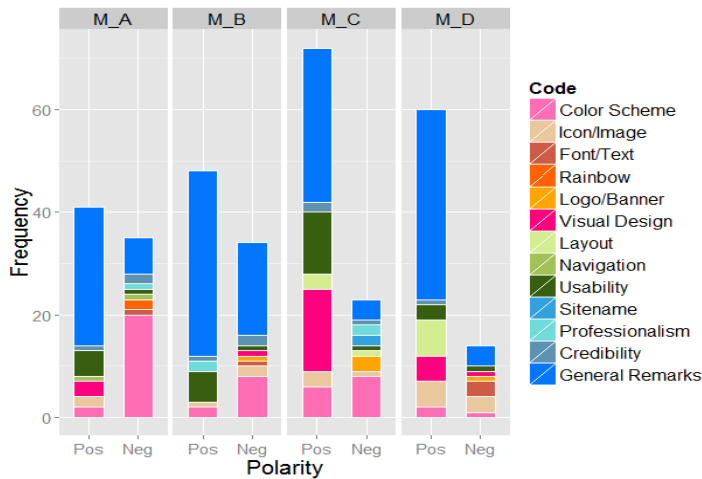
Figure 3. Principal components for comment files

4.3 Sentiment Analysis

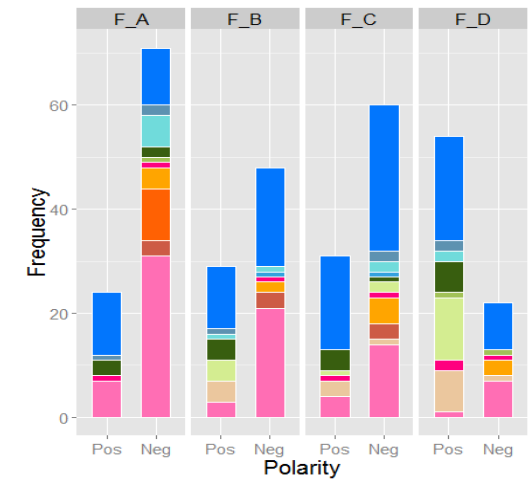
We carried out a sentiment analysis on each of the eight comment files in order to gain insight into: 1) what the first two principal components might represent; 2) what participants were most concerned with in each of the four UIs. For example, *what design themes run through the comment files in each and all clusters?* QDA Miner Lite [17] was used to manually code the comments into 13 subthemes. These subthemes were further categorized into three broader themes as shown in Table 3. *Visual design*, according to [5], refers to the aesthetics of a website coupled with its emotional appeal and balance. This may be expressed through sensory design elements such as colors, shapes, font type or multimedia *Usability*, according to our coding criteria, refers to convenience of use, ease of use, easy orientation and easy navigation of a website [34]. This can be expressed through layout, position and orientation of elements, such as buttons, texts, images, etc. *General remarks* refers to general comments relating to the site's name, professionalism and credibility. Lastly, general comments on visual design and usability as well as remarks, which do not fall under any of the sub-themes, are coded using the broader theme names, while very specific comments are coded using the subthemes names.

4.3.1 Positive vs. Negative Sentiments

Figure 4 shows a plot of the positive and negative sentiments for both genders under 13 identified subthemes which run through all eight comment files. Overall, the male comments (Figure 4a) are characterized by more positive and less negative sentiments than the female comments (Figure 4b). For example, the files containing all the male participants' comments on artefact A (M_A) comprises 42 positive and 35 negative sentiments, denoted as (+42, -35), for brevity. In contrast, F_A comprises (+24, -72) sentiments. Similarly, M_C comprises (+72, -23) sentiments, while F_C comprises (+31, -60) sentiments. Figure 4 also provides insight into some likely characteristics that defined the clusters in Figure 3. It reveals that M_A, M_B, M_C and M_D (files with male's comments) are more positive than their respective female counterparts (F_A, F_B, F_C and F_D). This seems to account for the location of M_A, M_B, M_C and M_D below the hypothetical line, $y=0$, and F_A, F_B, F_C, and F_D above it. As a result, the principal component B in Figure 3 may be regarded as a measure of sentiment polarity with files (M_A, M_B, M_C M_D, F_D) below the line, $y=0.25$, indicating more-positive-than-negative sentiments and those above it (F_A, F_B and F_C) indicating more-negative-than-positive sentiments.



(a) Male



(b) Female

Figure 4. Gender-based sentiments for all four UI

Table 3. Broad and subthemes in comment files

Broad theme	Sub-themes
Visual design	Color scheme, icon/image, font/text, rainbow theme and logo/banner
Usability	Layout and navigation
General remarks	Site name, professionalism and credibility

4.3.2 General vs. Specific Remarks

Figure 4 shows that, overall, the male group gave more general remarks in their response to the web designs than the female group. This is evident in the broader blue (general remark) band, green (usability) band and red (visual design) band in the male than the female bar chart. For example, regarding B, the male group made 54 general remarks (+36, -18), while the female group made about 32 general remarks (+12, -19). Similarly, regarding D, the male group made 41 general remarks (+37, -4), while the female group made about 30 general remarks (+20, -9). Typical examples of general remarks include: *Everything about it appeals to me* (M_A), *very attractive* (F_A); *the design is really interesting* (M_B), *very boring but less tacky* (F_B); *the page looks blurred* (M_C); *actually the mobile page is fascinating* (F_C). One would have expected females to record more general remarks than males given the fact that the former provided more comments (see Table 2). In contrast, the female group gave more specific remarks than the male group (see subsection 4.4.2). For example, in all four UIs, females used the specific word *color* and related words (e.g. blue, gray, black, white, etc.) much more than males as indicated by the broader color scheme band in the female bar chart. They also noticed and made specific reference to the rainbow color scheme used in A in their comments more than the males as shown in the broader rainbow band in the female bar chart. This supports the theory that females are more visually discerning [15] and more specific in decoding nonverbal cues [29] than males.

4.4 Documents Word Clouding

We combined all the comment files of each gender into one single file and carried out word clouding (minimum frequency of words = 5, scale is 5 to 1) on it in order to gain insight into what specific UI design elements participants were most concerned with, choice of words used, and how they vary across gender. Figure 5 shows the word clouds for both gender. It is discussed in the next subsections.

4.4.1 Visual Design vs. Usability Concerns

The most prominent theme that runs through the comment files for both group is visual design or look and feel (as evident in the boldness of such words as *color*, *look*, *nice*, *appeal*, *visual*, *attractive*, etc.). This resonated more with the FG than the MG (as evident in the bolder words, such as *look* and *image* in the female word cloud and Fig. 4). In particular, *color* turns out to be the overarching concern for both genders. This suggests that the color scheme chosen in the design of a mobile website is critical to its success or failure. The second most prominent theme is usability (as evident in the high occurrence of such words as *use*, *easy*, *simple*, *interesting*, *navigate*, etc.). This resonated more with the MG than the FG (as evident in the bolder words, such as *easy*, *use* and *navigate*, in the male word cloud).

4.4.2 Abstract/Generic vs. Concrete/Specific Words

As shown in Figure 5, males and females use different choices of words in expressing their reactions to the UI designs. Male tended to use more abstract and generic words, while females tended to use more concrete and specific words. By generic words we mean general remarks that do not refer to any specific part or element of the UI under assessment, e.g., *color*, *logo*, *icon*, *layout*, *background* etc. For example, apart from *color*, the next predominant visual design-related noun term in the male cloud is *design* (abstract and generic), while in the female cloud is *image* (concrete and specific). All in all, the graphical design-specific elements that made it into the male cloud are *color*, *image*, *icon*, *logo*, *background* and *blue*, while those that made it to the female cloud are *color*, *image*, *icon*, *logo*, *header*, *layout*, *font*, *buttons*, *rainbow*, *blue*, *gray* and *ranch* (representing G-Ranch).

4.4.3 Interface Commendation vs. Condemnation

Males tended to be more impressed with all four UIs than the females, and thus expressed more positive sentiments than the females. This is evident in such choice of words as *good*, *easy*, *attractive* and *interesting*, *simple*, *nice*, which are more frequent and prominent in the male word cloud. This finding is also evident in Figure 4a where males have more positive than negative sentiments. On the other hand, the female group tended to be more critical of the UIs than the male group as evident in choice of words such as *boring*, *ugly*, *lack* and *less*, which are present in the female cloud but not in the male cloud, which contains only two negative words (poor and dull). This is also evident in Figure 4 and Figure 6, which shows the overall sentiments of each gender. The overall sentiments expressed by females regarding the four interfaces, except D, is negative. This contrasts the overall sentiment expressed by males, which is positive. This qualitative result supports prior findings stating that females are more critical of websites than males [10], [8], [11], [30], [18]. Moreover, Figure 6 shows that the global overall sentiments for D, C and B is positive (with B almost having a zero value), while A is negative.

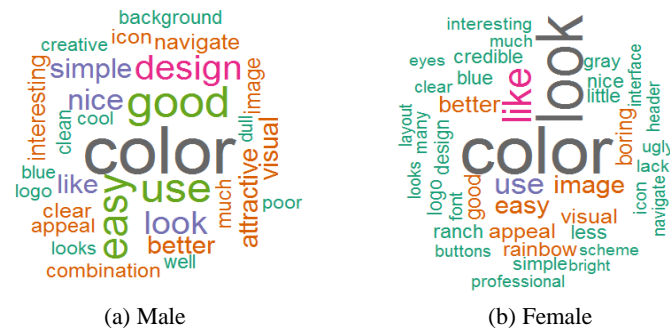


Figure 5. Word cloud for comments on all four webpages

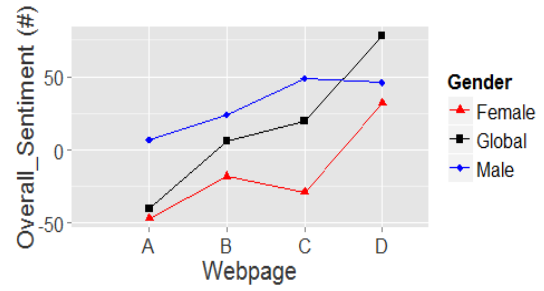


Figure 6. Overall (average) sentiment of UIs

4.5 Statistical Analysis of Credibility Scores

In addition to the qualitative analysis, we plotted the credibility rating and ranking of the four UIs by participants and performed a statistical analysis on them to confirm our qualitative findings. Figure 7 shows the plot of both measures on a 0-to-100% scale for both groups. Again, just as we have seen before, males rated all four interfaces in terms of credibility higher than the females, which confirms our findings in Section 4.3 and Section 4.4. We also see that as we move from one interface to another, participants' perception of credibility in terms of rating and ranking changed, with D being the best and A being the worse for both genders. While males rated all four interfaces more favorably (higher) than females, females ranked D, C and B higher than males. This was only possible because the ranking is forced; as a result, females' dislike and critical condemnation of A paved the way for the other three interfaces to rank higher.

4.5.1 Verification of H4: Between-group Analysis

Given that our data did not meet the normality requirement, we carried out the non-parametric Kruskal-Wallis rank (one-way ANOVA) test between the respective male and female credibility rating and credibility ranking scores of the four UIs to verify H4: *Females will be more critical in their judgment of the mobile websites than males*. Table 4 shows the result. First, with respect to the credibility rating, the test shows that there is a significant difference between the two groups. The group difference regarding A, B and C is significant at $p < 0.0000$, while that regarding D at $p < 0.0001$. This highlights how the two groups differ in their credibility perception, with the females being more critical than the males, as we have found before in the qualitative analysis (see Fig. 4). Second, with respect to credibility ranking, only the group difference regarding A is statistically significant at $p < 0.001$. A possible explanation for this, unlike the rating where the group difference is significant with regard to all four UIs, is that the ranking is forced and constrained between 1 and 4. There is a limited range of numbers to choose from in ranking compared to rating with a wider range of 1 to 7. However, the highly significant group difference in credibility ranking with regard to A highlights how much the male and female groups differ in the perception of A, as seen in the qualitative result (see Fig. 4). It indicates that females were completely disapproved the color scheme used in A.

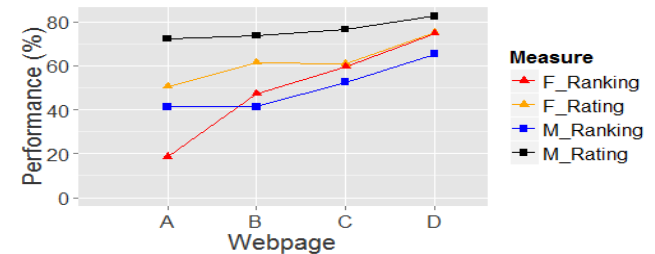


Figure 7. Credibility rating and ranking of UIs

Finally, based on the highly significant group difference regarding the four UIs' credibility rating at most at $p < 0.0001$ (Table 4), with females scoring lower (Fig. 7) and providing more negative than positive comments (Fig. 4), the fourth hypothesis (H4) is validated.

4.5.2 Verification of H1: Within-group Analysis

Table 5 shows the result of the non-parametric Friedman test and Nemenyi post-hoc pairwise comparison test [27] we carried out to verify our first hypothesis (H1: users' perception of credibility of mobile websites changes as the UI designs are modified). In the within-group analysis, with respect to pairwise significance, there is a correspondence between the credibility *rating* and credibility *ranking* measures for both groups. In other words, for each group, it is either both the *rating* and *ranking* pairwise comparison tests for a given UI transformation are significant ($p < 0.05$) or they are not significant (n.s). For example, for the MG, the results regarding *rating* for UI transformations, $A \rightarrow D$, $B \rightarrow D$, and $C \rightarrow D$, are significant at $p < 0.01$ (at most), and so are the corresponding results regarding *ranking* significant at $p < 0.05$ (at most). Similarly, for the FG, the results regarding *rating* for UI transformations, $A \rightarrow B$, $A \rightarrow C$, $A \rightarrow D$, $B \rightarrow D$, and $C \rightarrow D$, are significant at $p < 0.01$ (at most), and so are the corresponding results regarding *ranking* at $p < 0.05$ (at most). On the other hand, for the MG, the results regarding *rating* for UI transformations, $A \rightarrow B$, $A \rightarrow C$, and $B \rightarrow C$, are not significant, and so are those regarding *ranking* for these UI transformations. Similarly, for the FG, the result regarding *rating* for transformation $B \rightarrow C$ is not significant, and so is the result regarding *ranking* not significant. Therefore, since about 66% (16 out of 24) of the within-group (pairwise) comparison tests are significant at $p < 0.05$ (at most), we conclude that, to a great degree, our first hypothesis (H1) is supported.

4.5.3 Verification of H3/H4: Within-group Analysis

As shown in Table 4 and Figure 7, we see that D was rated and ranked as the best by both groups. The MG rated and ranked D 82.43% and 65.34% respectively, while the FG rated and ranked D 75.18% and 74.63% respectively. Based on the within-group (pairwise) analysis result shown in Table 5 and discussed in subsection 4.5.2, between the scores of D and each of the other three UIs, there is a significant difference at $p < 0.05$, with D being rated and ranked the highest in all 12 cases. Therefore, the third hypothesis (H3: *D will be judged as the best by both genders*) is validated. On the other hand, as shown in Table 5, given that the pairwise comparison ($A \rightarrow B$, $A \rightarrow C$ and $A \rightarrow D$) with respect to credibility rating and ranking for the FG is significant at $p < 0.0000$, the fourth hypothesis (H4: *A will be judged as the worst by both genders*) is supported for the FG. However, for the MG, except for $A \rightarrow D$, we see that the pairwise comparison ($A \rightarrow B$ and $A \rightarrow C$) with respect to credibility rating and ranking is not significant. Therefore, H4, for the MG, is not supported. A possible explanation for this is that the participants in the MG were, overall, liberal and not too critical in rating and ranking A and B which we perceived as low-level designs. Thus, we see a situation where there is no significant difference in the respective rating and ranking between A and B, and between A and C for this group.

4.5.4 Verification of H2: Sentiment Analysis

We used qualitative measure to verify the second hypothesis (H2: *Users will be more concerned about visual than navigational design elements in judging the mobile websites*). Figure 4 shows that users were more concerned about visual design (aesthetics) than navigational design (usability) when assessing the mobile websites. As shown in the bar chart, the visual design bars (red-like band) are broader than the usability bars (green-like band) for all four UIs. Thus, from a qualitative standpoint, H2 is confirmed.

Table 4. Kruskal-Wallis rank test between credibility Scores

Webpage	Credibility	Male	Female	Sig
A	Rating	72.07	50.62	$P < 0.0000$
	Ranking	41.22	18.66	$P < 0.0010$
B	Rating	73.76	61.44	$P < 0.0000$
	Ranking	41.22	47.26	$P = 0.0690$
C	Rating	76.58	60.70	$P < 0.0000$
	Ranking	52.25	59.45	$P = 0.0560$
D	Rating	82.43	75.18	$P < 0.0001$
	Ranking	65.34	74.63	n.s

Table 5. Friedman/Nemenyi pairwise credibility post-hoc test

Credibility	Gr	Cmp	Score1	Score2	Sig
Rating	M	A→B	72.07	73.76	n.s
		A→C	72.07	76.58	n.s
		A→D	72.07	82.43	$P < 0.0000$
		B→C	73.76	76.58	n.s
		B→D	73.76	82.43	$p < 0.0000$
		C→D	76.58	82.43	$p < 0.0100$
	F	A→B	50.62	61.44	$p < 0.0010$
		A→C	50.62	60.70	$p < 0.0100$
		A→D	50.62	75.18	$p < 0.0000$
		B→C	61.44	60.70	n.s
		B→D	61.44	75.18	$P < 0.0000$
		C→D	60.70	75.18	$p < 0.0000$
Ranking	M	A→B	41.22	41.22	n.s
		A→C	41.22	52.25	n.s
		A→D	41.22	65.34	$p < 0.0000$
		B→C	41.22	52.25	n.s
		B→D	41.22	65.34	$p < 0.0000$
		C→D	52.25	65.34	$p < 0.0500$
	F	A→B	18.66	47.26	$p < 0.0000$
		A→C	18.66	59.45	$p < 0.0000$
		A→D	18.66	74.63	$p < 0.0000$
		B→C	47.26	59.45	n.s
		B→D	47.26	74.63	$p < 0.0000$
		C→D	59.45	74.63	$p < 0.0500$

5. DISCUSSION

To synthesize our findings, we created a model to visualize and understand how the modification of the UI elements affected the credibility perception of the two groups (Figure 8). The model summarizes both the qualitative and quantitative findings. The corners of the rectangle, A, B, C, D, represent the four web designs, the blue arrows between each pair represent the UI transformations. The red and black colors represent the female and male groups respectively. The arrows indicate qualitative measure, where the upward direction indicates positive and the downward – negative overall sentiments. The lengths of these arrows represent the number of overall sentiments elicited by the respective UIs (see Figure 6). The pairs of values in brackets indicate the quantitative measures, where (v1, v2) represent credibility *rating* and credibility *ranking* scores respectively, and the asterisk (*) symbol between each male group score and each female group score indicates statistical significance when compared. The “+” sign between each pair of UIs indicates a significant increase in the perception of credibility (positive effect) when the UI is transformed from one design to another, while the “0” sign indicates no effect. These signs are based on the significance test results shown in Table 5.

5.1 UI Transformation Effect

Except from B to C, for the most part, we notice that as we transition from the low-level to the high-level designs, the perception and judgment of the UIs improve for both genders, with the FG recording more positives than the MG. The difference between the groups appears for transitions from A to B and A to C where there is a positive effect for FG and zero effect for MG.

5.1.1 Positive Effect UI Transformation

For the MG, we find that the UI transformations from any of the designs A, B, C to design D that resulted in a significant change (positive effect) in perception. For the FG, we find that all six UI transformations, except B → C, resulted in a positive effect in perception. The fact that there are more positive effects for the FG than the MG indicates that the FG's responses to the four UIs vary much more than the MG's, as we saw in the clustering of the comment files in Fig. 3. The FG's overall sentiments began from negative (at A), remained negative (at B and C) and only became positive (at D). On the other hand, the MG's overall sentiments began from neutral or slightly positive (at A), and remained positive all through (at B, C and D). This suggests that, in practice, females would be more responsive or sensitive to UI upgrades in a mobile website than males (see Fig. 6 and Fig. 7). Thus, operators of websites should ensure such upgrades are for the better; otherwise, the credibility of the sites, especially among females, may decline.

5.1.2 UI Transformation with Different Effect

The two groups differ regarding A → B and A → C. The UI transformation had a zero effect on the MG, but positive effect on the FG. An explanation for this difference can be found in the sentiment analysis: females tend to be more sensitive to and critical of the use of color in general. In particular, the rainbow color scheme was highly penalized by the FG. However, upon “improving” the multicolor scheme (A) to the gray (B) and blue (C) color schemes, which they perceived as more professional and appealing respectively, they toned down their criticism, as evident in the better credibility rating and ranking scores and the less negative overall sentiments about B and C. However, the MG was not as critical of A at first as the FG, as seen in the credibility rating scores of 51 and 72 respectively. Thus, we see very close credibility rating scores between A and B, and between A and C, for the MG. For example, the credibility rating difference between A and B is 2 and that between A and C is 5, which are not significant (“0”). In contrast, for the FG, the credibility rating differences between A and B, and between A and C are high: that between A and B is 10 and that between A and C is 10, which are both significant (“+”).

5.1.3 Zero-Effect UI Transformation for both Groups

The zero-effect transformation between B (a low-level design) and C (a high-level design) for both genders is also evident in the least change of global overall sentiment (see Figure 6 also). This could be explained based on participants' comments on the UIs. They generally tended to view the gray color theme of B as a professional theme and thus were not too critical about it (relative to C). This suggests that, in practice, if every other UI design characteristic is okay, users may be less bothered by a gray color scheme than by another color scheme, which may be perceived as unappealing and unprofessional, e.g. A. However, from the sentimental analysis, it seems that users would prefer a mobile site both professional and colorful, as they want to see some level of color, which appeals to their visual sense. Moreover, it is noteworthy how the FG assessed C qualitatively (see Fig. 4b), especially regarding its color scheme³. One would have expected that its overall assessment would be positive given that blue is a common theme used by many popular websites (e.g., Facebook) and mobile websites (e.g., banking).

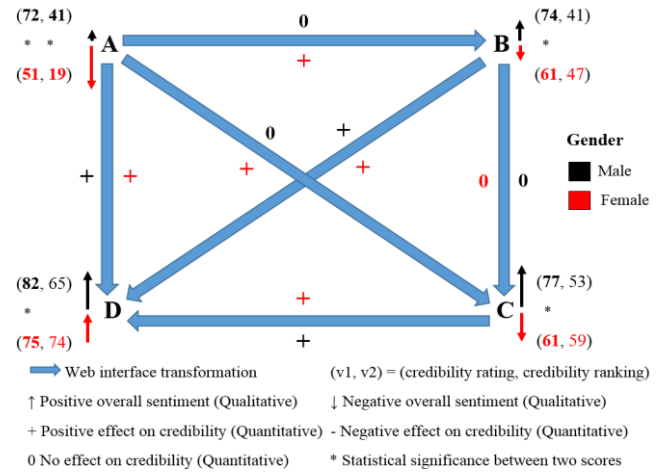


Figure 8. User Model for understanding UI perception

However, the FG's overall assessment of C turned out to be negative and even worse than B which employed a gray color scheme, described by some female participants as “boring” and “bland”. The sentiment analysis showed that the FG were displeased with C's blue color scheme. Thus C and B ended up having the same credibility rating of 61. The reason may be due to C's unicolor scheme. Some saw this as “boring” and “hard on the eyes”. Some typical comments include: 1) *The blue is horrible to look at (hard on the eyes, physically)...* 2) *The blue and blocks seems very boring, and does not make me want to use the site...* 3) *The website does not appeal to me because of the color combination.* However, on modifying C's layout from list to grid (D), the FG seemed not to be concerned about the all-blue color scheme any longer. They tended to focus on the “ease of use” of D, which seemed to increase its appeal to them. As a result, their overall sentiment shifted positive, with the credibility rating and ranking increasing from 61 and 59 (at C) to 75 and 74 (at D) respectively: an indication of the halo effect [33]. A possible explanation for this change in perception, which also explains why D is the most preferred by both genders, is evident in participants' comments. People generally are so used to the iPhone (grid) layout for mobile apps that they expect to see it also in mobile web design. Besides, most smart phones use this layout for placing their apps on the screen, as this has become a *de facto* standard. Therefore, the participants found it more credible, professional, and aesthetic. This suggests that, in practice, due to its relative ease of use, the grid layout should be given priority by designers, especially when all the mobile web application's content can fit into one screen.

In a nutshell, the following, based on the sentimental analysis, are noteworthy for the adaptation of mobile websites based on gender:

1. Females care more about visual design than males.
2. Females care more about the professional use of images, fonts and a multicolor theme in mobile sites than males.
3. Males care more about usability features than females.

Moreover, Table 5 shows a set of mobile website design guidelines, informed by participants' comments. It includes key features, justifications and snapshots of participants' comments. The first key feature emphasizes the need to use a color scheme that is both appealing and professional, as improper use of colors may cause users to doubt the site's credibility. The second key feature focuses on the site name and logo. Users want to see a site whose name and logo, as much as possible, reflect the products and services of the site. For example, given that the hypothetical website we presented

Table 5. Empirically backed feature set of guidelines for the design of mobile websites

	Expected Key Features	Justification	Comments
1.	The color scheme or theme used in the mobile website design should be appealing and/or professional.	Most participants felt the color scheme used in A was unprofessional and amateurish, while that used in B, though might be more credible, was old-fashioned and boring.	<i>More colors make it more fascinating, but it still looks sort of amateurish. Maybe it's the font type? (F_A). It looks nice but the grey scale is very boring (F_A).</i>
2.	Site name and logo should be as intuitive as possible to give users a quick insight into what products and services the site offers.	Some participants felt the logo/banner of the hypothetical website was not good enough, the name did not reflect its services. A few participants suggested a green theme best suited the website given the name "G-Ranch", which reflected tourism.	<i>"G" ranch makes me think its not a legit, I would use something else. Company logo or some info or anything would make it look less fake (F_C). Name plays into the credibility level too. G-Ranch doesn't make me think it's a real place because it sounds too simple and made up. (M_C).</i>
3.	Icon menu should be preferred to color-bar menu if website content (items) is to be presented using a list layout.	Most participants preferred D, B and C (which used icon menu) to A (which used color bars). They believe this is more intuitive in the presentation of contents.	<i>This web page is outstanding and the pictorial illustrations are self-explanatory such that even if one cannot read English, the symbols will guide such ones. Its excellent job done. (M_D).</i>
4.	Grid layout should be preferred to list layout. Or better still, provide users with a layout option.	Most participants (both males and females) preferred D to C, as they felt it was more navigable and less prone to error in the course of clicking on an item.	<i>The menu are well spelt out and are easy to navigate (M_D). The buttons would probably be easy to press with my stubby fingers without accidentally hitting the wrong one (F_D).</i>
5.	Help should be provided to assist non-expert users in navigating the site easily.	Some participants expected to see a help feature where supportive information on the usage of the site could be sought.	<i>There should be a "Help" button to understand the app better. An intro to the app would have increased its credibility (F_B).</i>
6.	Search feature should be provided users to help users find information easily.	Some participants requested a search box where they could type in and search for information they wanted	<i>I don't see any place where I can type in a search for what I want (F_C). It's annoying that there is no option to search (F_C).</i>
7.	Language option or translation feature should be provided if possible.	A couple of participants suggested language option be provided to enable the non-default language speakers to use the site as well.	<i>Great layout but language option should be added for non-English speaking countries (F_B).</i>
8.	Extraneous allusions should be avoided in the website design, branding and presentation of information.	A couple of participants were put off by the rainbow menu used in A, as they deemed this inappropriate, unprofessional and unconnected to the site.	<i>Looking more credible but the bright rainbow colors are not necessary (F_A). The colorful rainbow does not look professional (F_A). Is this App for Kids? Rainbow Menu (M_A).</i>
9.	Users should be allowed to access site, at least its basic features, without having to sign up or sign in.	A couple of participants were bothered with the sign-in button below the site banner, as they felt reluctant to sign-in before having full access to what the site had to offered.	<i>It looks as if there is a sign-in option, which would worry me - do I need to waste time entering my info? (F_C).</i>

relates to tourism, some participants expected the site to have a green theme, which reflected the fictitious name "G-Ranch" and services, such as *golf*. Other key features, which they expected to see on the site, include *search*, *help*, *language option*, etc. Finally, thanks to the useful information on users' expectations, gathered from the sentiment analysis, we recommend that in the design of a mobile website, as a way of formative evaluation, potential users (with a gender balance) should be involved in the design process in order to gather useful qualitative information on users' needs [36].

5.2 Limitation

One of the limitations of our study is the order in which the mobile webpages (C, A, B and D) were presented to participants. It would have been better if we had been able to randomize the order.

6. CONCLUSION

We have presented the findings of a mixed-method study on users' perceptions of four systematically modified mobile websites and the role gender plays based on a mixed sample of 285 subjects from North America, Africa and Asia. We showed both quantitatively and qualitatively that the design of a mobile website affects the

perception of its credibility, with females being more critical, responsive and sensitive to UI changes than males, and the grid-layout preferred to list-layout design by both genders. Our findings, by implication, reiterates the need for mobile website vendors to provide users with customizable mobile websites, which they can tailor to their thematic and layout preferences. More important, it would better pay off if vendors can infer the gender and preferences of their visitors/users and personalize such essential look and feel of the UI as theme and layout, accordingly, instead of having users do it themselves. Our contributions to knowledge are in two fold. First, regarding mobile web design: 1) we confirmed the existing theory that holds that gender difference exists in the perception of mobile sites, with females being more critical in their judgement; 2) we showed that as UI design characteristics change, users' perceptions change also, with females being more sensitive and responsive to those changes; 3) we showed that both genders prefer the grid to the list layout; and 4) we presented an empirically backed set of guidelines for the design of mobile websites. Second, regarding mobile web adaptation, we presented important findings that can inform a gender-based site adaptation, e.g., females liking ample colors professionally used alongside images and fonts.

7. REFERENCES

- [1] Arcand, M. and Nantel, J. 2005. Gender differences in processing information: Implications for online search patterns and task performance. HEC Montréal.
- [2] Bergkvist, L. and Rossiter, J.R. 2007. The Predictive Validity of Multiple-Item Versus Single-Item Measures of the Same Constructs. *Journal of Marketing Research*. 44, 2 (2007), 175–184.
- [3] Boiano, S. et al. 2006. Gender Issues in Hci Design for Web Access. *Advances in Universal Web Design and Evaluation: Research, Trends and Opportunities*. S. Kurniawan and P. Zaphiris, eds. Idea Group Publishing. 116–153.
- [4] Burstein, B. et al. 1980. Sex differences in cognitive functioning - evidence, determinants, implications. *Human Development*. 23, (1980), 289–313 ST – Sex differences in cognitive function.
- [5] Cyr, D. et al. 2006. Design aesthetics leading to m-loyalty in mobile commerce. *Information and Management*. 43, 8 (2006), 950–963.
- [6] Cyr, D. 2009. Gender and website design across cultures. *17th European Conference on Information Systems*. (2009), 279–291.
- [7] Cyr, D. et al. 2008. Web site design, trust, satisfaction and e-loyalty: the Indian experience. *Online Information Review*. 32, 6 (2008), 773–790.
- [8] Cyr, D. and Bonanni, C. 2005. Gender and website design in e-business. *International Journal of Electronic Business*. 3, 6 (2005), 565.
- [9] Darley, W.K. and Smith, R.E. 1995. Gender differences in information processing strategies: An empirical test of the selectivity model in advertising response. *Journal of Advertising*. 24, 1 (1995), 41–56.
- [10] Ferebee, S. 2008. The influence of gender and involvement level on the perceived credibility of web sites. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. 5033 LNCS, (2008), 279–282.
- [11] Flanagan, A.J. and Metzger, M.J. 2003. The perceived credibility of personal Web page information as influenced by the sex of the source. *Computers in Human Behavior*. 19, 6 (2003), 683–701.
- [12] Fogg, B.J. et al. 2003. How do users evaluate the credibility of Web sites?: a study with over 2,500 participants. *Proceedings of the 2003 conference on Designing for user experiences* (2003), 1–15.
- [13] Fogg, B.J. et al. 2001. What makes Web sites credible?: a report on a large quantitative study. *Proceedings of the SIGCHI conference on Human factors in computing systems* (2001), 61–68.
- [14] Geary, D.C. 1996. Sexual selection and sex differences in mathematical abilities. *Behavioral and Brain Sciences*. 19, 02 (1996), 229.
- [15] Holbrook, M.B. 1986. Aims, Concepts, and Methods for the Representation of Individual Differences in Esthetic Responses to Design Features. *Journal of Consumer Research*. 13, 3 (1986), 337.
- [16] Hyde, J.S. and Linn, M.C. 1988. Gender differences in verbal ability: A meta-analysis. *Psychological Bulletin*. 104, 1 (1988), 53–69.
- [17] Lewis, R.B. and Maas, S.M. 2007. QDA Miner 2.0: Mixed-Model Qualitative Data Analysis Software. *Field Methods*. 19, (2007), 87–108.
- [18] Li, N. et al. 2001. Cross-cultural comparison of women students' attitudes toward the Internet and usage: China and the United Kingdom. *CyberPsychology & Behavior*. 4, 3 (2001), 415–426.
- [19] Li, Y.M. and Yeh, Y.S. 2010. Increasing trust in mobile commerce through design aesthetics. *Computers in Human Behavior*. 26, 4 (2010), 673–684.
- [20] Lindgaard, G. et al. 2006. Attention web designers: You have 50 milliseconds to make a good first impression! *Behaviour & information technology*. 25, 2 (2006), 115–126.
- [21] Liu, C. et al. The Influence of HCI Design Aesthetics on Website Creditability – Using an Online Banking Website Interfaces as an Example. 1–20.
- [22] Louann Brizendine 2006. *The Female Brain*. Morgan Road, New York.
- [23] Lu, Y. and Rastrick, K. 2014. Impacts of Website Design on the Adoption Intention of Mobile Commerce: Gender as a Moderator. *New Zealand Journal of Applied Business Research*. 12, 2 (2014), 51–69.
- [24] Meyers-Levy, J. 1989. Gender Differences in Information Processing: A Selectivity Interpretation. *Cognitive and Affective Responses to Advertising*. (1989), 219–260.
- [25] Meyers-Levy, J. and Maheswaran, D. 1991. Exploring Differences in Males' and Females' Processing Strategies. *Journal of Consumer Research*. 18, 1 (1991), 63–70.
- [26] Oyibo, K. et al. 2016. An Empirical Analysis of the Perception of Mobile Website Interfaces and the Influence of Culture. *Persuasive Technology 2016 Workshop one Personalization in Persuasive Technology* (Salzburg, Australia [Accepted], 2016), 1–12.
- [27] Pohlert, T. 2016. The Pairwise Multiple Comparison of Mean Ranks Package (PMCMR). *R package*. (2016), 1–9.
- [28] Robins, D. and Holmes, J. 2008. Aesthetics and credibility in web site design. *Information Processing & Management*. 44, 1 (2008), 386–399.
- [29] Rosenthal, R. and DePaulo, B.M. 1979. Sex Differences in Accommodation in Nonverbal Communication. *Skill in Nonverbal Communication: Individual Differences*. R. Rosenthal, ed. Oelgeschlager, Gunn and Hain. 68–103.
- [30] Schumacher, P. and Morahan-Martin, J. 2001. Gender, Internet and computer attitudes and experiences. *Computers in human behavior*. 17, 1 (2001), 95–110.
- [31] Setterstrom, S. 2010. *Assessing credibility and aesthetic perception across different exposure times on a health care information website*. Iowa State University.
- [32] Simon, S.J. 2001. The impact of culture and gender on web sites: an empirical study. *SIGMIS Database*. 32, 1 (2001), 18–37.
- [33] Soper, D.S. 2014. User Interface Design and the Halo

Effect : Some Preliminary Evidence. *Twentieth Americas Conference on Information Systems, Savannah* (2014), 1–11.

- [34] Sutcliffe, A. and De Angeli, A. 2005. Assessing interaction styles in web user interfaces. *Human-Computer Interaction-INTERACT 2005* (2005), 405–417.
- [35] The fastest-growing mobile phone markets barely use apps: 2015. [http://qz.com/466089/the-fastest-growing-](http://qz.com/466089/the-fastest-growing-mobile-phone-markets-barely-use-apps/)

[mobile-phone-markets-barely-use-apps/](http://qz.com/466089/the-fastest-growing-mobile-phone-markets-barely-use-apps/). Accessed: 2016-03-03.

- [36] Kristine Faulkner 2000. *Usability Engineering*. Palgrave.
- [37] 2014. Digital Impact on In-Store Shopping: Research Debunks Common Myths. *Google*. (2014).